

Ammar Sidhu

(647-892-6096) | ammarsidhu@outlook.com | [linkedin/ammarsidhu/](https://www.linkedin.com/in/ammarsidhu/) | [github/ammarsidhu](https://github.com/ammarsidhu) | ammarsidhu18.github.io

EDUCATION

University of Toronto

Toronto, ON

Bachelor's of Science in Applied Statistics and Geographic Information Systems

Sept. 2018 - Apr. 2024

- **Minor** in Mathematical Sciences
- **Coursework:** Spatial Data Science, Regression Analysis, Bayesian Statistics, Information Processing, Space Time Data Analysis

EXPERIENCE

University of Toronto

Toronto, ON

Undergraduate Research Analyst

June 2022 - Aug. 2022

- Automated landuse/landcover (LULC) mapping of the Toronto and Region Conservation Authority (TRCA) with classification algorithms.
- Created a training dataset in **QGIS** by **sampling over 1000 pixels with stratified random sampling** based on **16 LULC classes**.
- Implemented **Sentinel-2 satellite** imagery's spectral, and visible bands (**10m spatial resolution**) as predictor data.
- Applied **cloud-masking** on the satellite imagery to filter any cloud-cover that was present in the TRCA during peak summer (June to August).
- Achieved a testing data **accuracy of 74%** across all LULC on the **Random Forest** model through leveraging **hyperparameter Tuning**.

Canada Post

Mississauga, ON

Data Entry

Apr. 2019 - Jan. 2021

- Assisted Canadian Border Services Agency with entering parcel information using excel spreadsheet automation formulas to sort mail.
- Resulted in a **25% increase** in mail distribution during the COVID-19 pandemic.

PORTFOLIO PROJECTS

Employee Salary Prediction Based on Credentials and Qualifications ([View](#))

- Worked with an employee salary dataset from a recruiting company containing **1,000,000 employees** across **8 individual credentials (features)**.
- Developed a **Salary Prediction Model** to help recruiters determine employee salaries by building regression models in Scikit-Learn.
- Obtained a **MSE of 93%** on the test dataset with the **Gradient Boosting Regressor** algorithm after hyperparameter tuning with **GridSearchCV**.

Can We Classify How a Nuclear Bomb was Deployed? ([View](#))

- Developed a **Nuclear Bomb Deployment Classifier** that identifies the deployment method of a nuclear bomb given features about the bomb.
- Trained and tested ensemble algorithms on **imbalanced multiclass data (4 classes)** with **Cost-Sensitive Learning** in Scikit-Learn.
- Achieved **92% accuracy** on the **Random Forest Classifier** after experimenting with **60 sets** of hyperparameters in **RandomizedSearchCV**.

The Influence of Location for House Price Predictions - Spatial Regression vs. Machine Learning ([View](#))

- Explored and mapped house pricing data for **187 census tracts** for the city of Hamilton by working with Canada's Census Tract data.
- Built a **House Price Predictor** via **Ridge Regression** with **MSE of \$57,070.99**, and **R² of 0.80** to help predict house prices by census tract.
- Improved prediction accuracy by **increasing the R² to 0.85** on **log-transformed data** by implementing the **Spatial Lag Regression model**.

Predicting Heart Disease in Patients ([View](#))

- Constructed a **Heart Disease Classifier** with linear and ensemble classifiers in Scikit-Learn on Boston Hospital's heart disease data.
- Achieved **92% testing data accuracy** on the **Random Forest Classifier** by implementing hyperparameter tuning with **GridSearchCV**.
- Determined by **feature importance** that thallium stress and number of major vessels colored by fluoroscopy are strong predictors of heart disease.

TECHNICAL SKILLS

Languages: Python, R, SQL, HTML; Familiar with JavaScript, CSS

Libraries & Frameworks: Pandas, Scikit-Learn, TensorFlow, NumPy, Seaborn, Matplotlib, Tidyverse, BeautifulSoup

Software: Google Earth Engine, ArcGIS Pro, QGIS, Git, Excel/Google Sheets; Familiar with GCP, Tableau, Hadoop